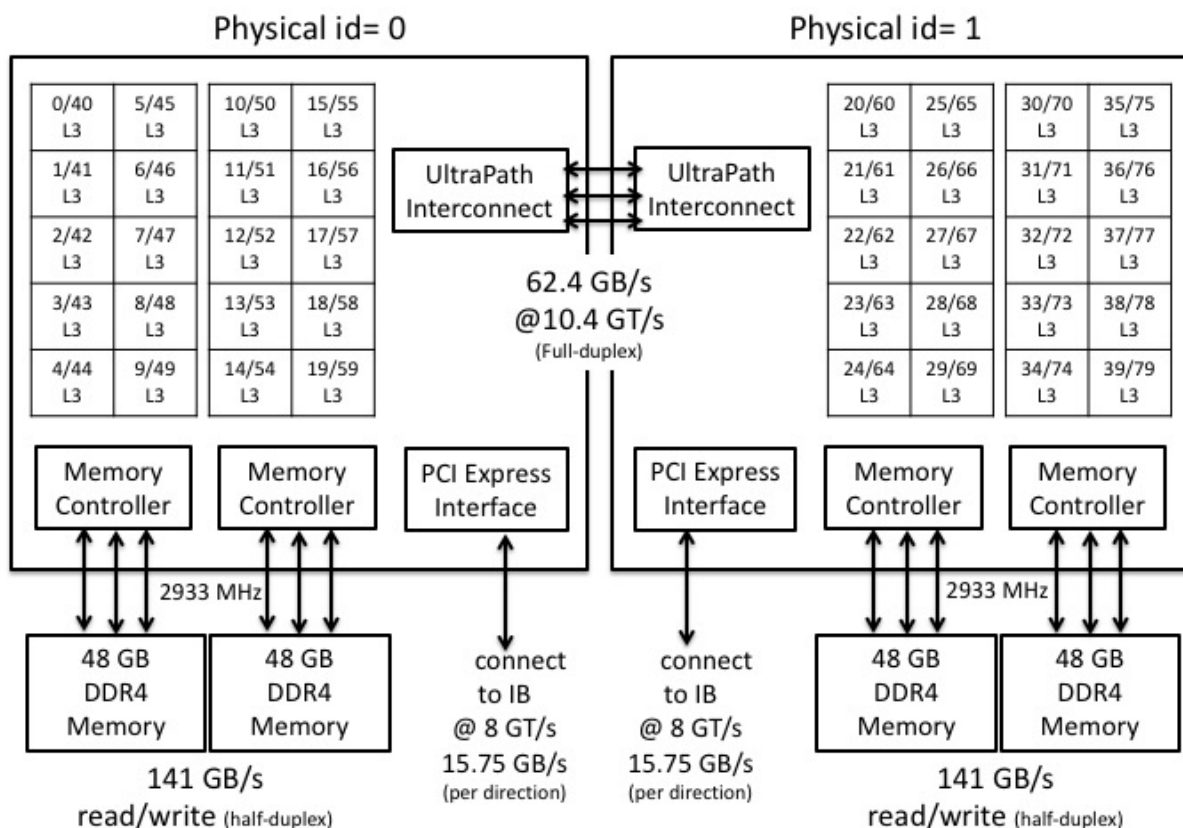# Cascade Lake Processors

Configuration of a Cascade Lake - SP Node

Notes:

- Each small square in the diagram represents a combination of a physical core and a L3 cache slice. For each physical core, there are two logical core labelings obtained from the **"processor"** entry of the `cat /proc/cpuinfo` output.
- Although Cascade Lake processors support up to three Intel Ultra Path Interconnect (UPI) links with a bandwidth of 62.4 GB/s, the HPE 8600 Saxon board used for the Cascade Lake nodes at NAS implements only two links, with a bandwidth of 41.6 GB/s.

The Intel Cascade Lake processor incorporated into the Aitken cluster is the 20-core Xeon Gold 6248 model. Its base clock speed is 2.5 GHz for non-AVX, 1.9 GHz for AVX2, and 1.6 GHz for AVX-512. It uses an enhanced 14-nanometer (nm) fabrication process and the Skylake microarchitecture with some optimization.

Each Aitken Cascade Lake node contains two Cascade Lake processors and uses a dual single-port 100 Gbits/s Enhanced Data Rate (EDR) Mezzanine card to connect outward (see Inter-node Network). The use of four 200 Gbits/s High Data Rate (HDR) switches per enclosure forms the MPI and I/O InfiniBand fabrics within the Aitken cluster. Connecting the Aitken cluster to the Pleiades filesystems relies on additional sets of HDR switches and cables.

Instruction Sets

In addition to the instruction sets SSE, SSE2, SSE3, Supplemental SSE3, SSE4.1, SSE4.2, AVX, AVX2, AVX-512, and AVX512[F,CD,BW,DQ,VL], which are available in its Skylake predecessor, Cascade Lake also includes the new AVX-512 Vector Neural Network Instructions (VNNI), which provide significant, more efficient deep-learning inference acceleration. Cascade Lake also introduces in-hardware mitigations for the Spectre and Meltdown security flaws.

With 512-bit floating-point vector registers and two floating-point functional units, each capable of Fused Multiply-Add (FMA), a Cascade Lake core can deliver 32 double-precision floating-point operations per cycle.

Use the Intel compiler flag `-xCORE-AVX512` for Skylake and Cascade Lake-SP specific optimizations. The optimization flag `-qopt-zmm-usage=high -xCORE-AVX512` may benefit floating-point heavy applications running on Skylake and Cascade Lake.

Tip: If you want a single executable that will run on any of the Aitken, Electra, and Pleiades processor types, with suitable optimization to be determined at run time, you can compile your application using the option:
`-O3 -ipo -axCORE-AVX512,CORE-AVX2 -xAVX`.

Hyperthreading

Hyperthreading is turned ON.

Turbo Boost

Turbo Boost is turned ON. Maximum Turbo Frequency is 3.90 GHz for non-AVX, and 3.8 GHz for AVX2 and AVX-512.

Cache Hierarchy

The cache hierarchy of Cascade Lake is as follows:

- L1 instruction cache: 32 KB, private to each core; 64 sets; 64 B/line; 8-way
- L1 data cache: 32 KB, private to each core; 64 sets; 64 B/line; 8-way; fastest latency: 4 cycles; 128 B/cycle load bandwidth; 64 B/cycle store bandwidth; write-back policy
- L2 cache: 1 MB, private to each core; 64 B/line; 16-way; fastest latency: 14 cycles; 64 B/cycle bandwidth to L1 cache; write-back policy
- L3 cache: shared non-inclusive 1.375 MB/core; total of 27.5 MB, shared by 20 cores in each socket; 2,048 sets; 64 B/line; fully associative; fastest latency: 50 - 70 cycles; write-back policy

Memory Subsystem

Like Skylake, there are two sub-NUMA clusters in each Cascade Lake socket, creating two localization domains. There are three memory channels per sub-NUMA cluster. Each channel can be connected with up to two memory DIMMs. For the Aitken Cascade Lake configuration, there is one 16-gigabyte (GB) dual rank DDR4 DIMM with error correcting code (ECC) support per channel. In total, the amount of memory is 48 GB per sub-NUMA cluster, 96 GB per socket, and 192 GB per node.

The speed of each memory channel is increased from 2,666 MHz in Skylake to 2,933 MHz in Cascade Lake. An 8-byte read or write can take place per cycle per channel. With a total of six memory channels, the total half-duplex memory bandwidth is approximately 141 GB/s per socket.

On-chip Interconnect

The on-chip architecture of Cascade Lake SP uses the same mesh layout as that of Skylakeâ   where the cores and L3 caches are organized in rows and columnsâ   instead of the ring architecture used by earlier Xeon processors.

Inter-socket Interconnect

The Cascade Lake processors support up to three UPI links. The configuration at NAS has two links enabled. The UPI runs at a speed of 10.4 gigatransfers per second (GT/s). Each link contains separate lanes for the two directions. The total full-duplex bandwidth is 62.4 GB/s with three links and 41.6 GB/s with two links.

Inter-node Network

Like the network subsystems of the Pleiades Haswell and Broadwell nodes, and the Electra Broadwell and Skylake nodes, each Cascade Lake node uses two PCI Express interfaces (one from each socket) to connect to the Aitken and Pleiades InfiniBand (IB) fabrics. The use of PCIe 3.0 16-lane (x16) links enables a maximum bandwidth of 15.75 GB/s for each direction.

Like the Electra Skylake nodes, the Aitken Cascade Lake inter-node network makes use of the 100 Gbits/s EDR technology. As shown below, one PCIe is connected to the ib0 fabric via a single-port, four-lane (4X) EDR host channel adapter (HCA), in a dual single-port EDR IB mezzanine card, with an effective bandwidth of 100 Gbits/s (that is, 12.5 GB/s) for each direction. The other PCIe is connected to the ib1 fabric via another single-port 4x EDR HCA in the same mezzanine card.

NAS Cascade Lake
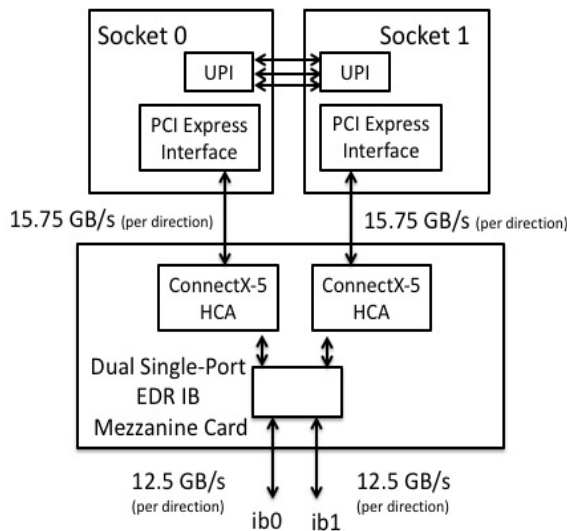
There are 1,152 Cascade Lake nodes in the Aitken cluster, which uses the liquid-cooled HPE 8600 system architecture. The nodes are partitioned as follows:

- Four E-cells (288 nodes per E-cell)
- Two compute racks and one cooling rack per E-cell (144 nodes per compute rack)
- Four enclosures (IRUs) per compute rack (36 nodes per enclosure)
- Nine compute trays (HPE XA730i "Saxon" blades) per enclosure (four nodes per tray)

The connection of the compute nodes relies on the use of four standard IB HDR switch blades per enclosure. Two of these switch blades facilitate the ib0 fabric, while the other two facilitate the ib1 fabric. Each switch blade has one 40-port ConnectX-6 ASIC with 18 ports connecting to the compute nodes, and 22 ports available for switch-to-switch links and connections to external targets. The connection of the 1,152 Cascade Lake nodes forms a 6-dimensional enhanced hypercube.

Notes:

- There are eight ASICs per enclosure in Electra Skylake, while there are four in Aitken Cascade Lake. This reduces a 1,152-node topology from seven dimensions in Electra Skylake to six dimensions in Aitken Cascade Lake.
- Only two out of the three UPI links are enabled.

References

- Intel Xeon Gold 6248 Processor
- Second Generation Intel Xeon Scalable Processors
- Intel 64 and IA-32 Architectures Optimization Reference Manual (Apr 2019 edition)
- HPE SGI 8600 QuickSpecs (Apr 15, 2019 edition)